Information Systems 122 (2024) 102367

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

ArZiGo: A recommendation system for scientific articles

Iratxe Pinedo, Mikel Larrañaga *, Ana Arruarte

Languages and Information Systems, University of the Basque Country UPV/EHU, Manuel Lardizabal Ibilbidea, 1, Donostia, 20018, Basque Country, Spain

ARTICLE INFO

Keywords:
Recommender system
Research articles
User profile
Content-based filtering
Collaborative filtering
Hybrid system

ABSTRACT

The large number of scientific publications around the world is increasing at a rate of approximately 4%–5% per year. This fact has resulted in the need for tools that deal with relevant and high-quality publications. To address this necessity, search and reference management tools that include some recommendation algorithms have been developed. However, many of these solutions are proprietary tools and the full potential of recommender systems is rarely exploited. There are some solutions which provide recommendations for specific domains, by using ad-hoc resources. Furthermore, some other systems do not consider any personalization strategy to generate the recommendations. This paper presents *ArZiGo*, a web-based full prototype system for the search, management, and recommendation of scientific articles, which feeds on the Semantic Scholar Open Research Corpus, a corpus that is growing continually with more than 190M papers from all fields of science so far. *ArZiGo* combines different recommendation approaches within a hybrid system, in a configurable way, to recommend those papers that best suit the preferences of the users. A group of 30 human experts has participated in the evaluation of 500 recommendations in 10 research areas, 7 of which belong to the area of Computer Science and 3 to the area of Medicine, obtaining quite satisfactory results. Besides the appropriateness of the articles recommended, the execution time of the implemented algorithms has also been analyzed.

1. Introduction

The online publication of research articles is continuously growing, with millions of documents scattered across the Internet, making it difficult for researchers to find publications suitable for their own interests. In this scenario, managing content of interest has become a challenge that had led to the development of different tools, ranging from search engines that allow interesting publications to be discovered (IEEE Xplore, ACM Digital Library…), reference management systems that facilitate the management and use of bibliographical references (EndNote, Mendeley, Zotero, Docear…), social networks that allow the creation of communities of researchers with common interests (ResearchGate, Academia.edu, ArXiv…) to recommender systems that allow articles of interest to be filtered (Mr. DLib). Although the development of these tools has made it possible to open the scientific world to new audiences, most of these systems correspond to private and proprietary initiatives.

Among the above-mentioned tools, Recommender Systems (RS) have achieved growing interest from the scientific community. Up to now, RSs have been successfully used in order to deal with the so-called paradox of choice, by helping users to find items they want in different domains. Although RSs originally arose in the domain of e-commerce, in recent years, their use and application has been extended to other domains such as education [1]. For example, RSs can be found in e-learning environments where many courses are offered [2], or when looking for the appropriate scholarly venue to send a publication to [3,4]. Currently, Research-Paper Recommender Systems (RP-RS) hold a prominent place, adopting and creating approaches and techniques that best suit their characteristics [5]. In practice, RPRSs mostly appear as a side complement within systems with a broader purpose [6].

Therefore, the development of a RPRS with an open and research-oriented perspective whose main objective is the generation of personalized scientific paper recommendations over the ongoing interests of users in a multidisciplinary and highly scalable environment is extremely necessary. ArZiGo (Artikulo Zientifikoen Gomendio-sistema; a Basque translation of Recommendation System for Scientific Articles), a hybrid research paper RS prototype, has been developed to fulfill this necessity.

In the review of the state-of-the-art in the area of RP-RSs presented in the following section, some drawbacks have been identified:

- the need of knowledge for the domain to provide recommendations,
- the lack of personalization in some recommendations,
- the use of offline evaluations in most of the works.

* Corresponding author.
E-mail addresses: iratxe.pinedo@ehu.eus (I. Pinedo), mikel.larranaga@ehu.eus (M. Larrañaga), a.arruarte@ehu.eus (A. Arruarte).

https://doi.org/10.1016/j.is.2024.102367
Received 1 June 2022; Received in revised form 10 November 2023; Accepted 4 February 2024
Available online 6 February 2024
0306-4379/

With *ArZiGo* (*Artikulo Zientifikoen Gomendio-sistema*; a Basque translation of *Recommendation System for Scientific Articles*), we will adapt approaches from content-based and collaborative filtering thanks to:

- its large and multidisciplinary catalogue of articles that is continuously growing with new publications,
- its interaction processing module, where the preferences of the users are identified and modeled considering their activities,
- its configurable and scalable design.

All of which will allow *ArZiGo* to:

- offer access to papers belonging to a wide range of domains and recommend, among these, the most appropriate ones according to the interests of the user,
- enable the users to manage the content of their interest, such as storing articles or following authors,
- generate personalized recommendations for users based on their interaction history,
- provide a configurable hybrid recommendation module, which makes it easy to incorporate new recommendation techniques.

The paper is structured as follows. Section 2 introduces the technologies that constitute the basis for the development of Research-Paper Recommender Systems (RP-RSs) and the most recent work in the domain. Section 3 presents the main characteristics of *ArZiGo* and describes each of its components. Section 4 deals with the evaluation methodology proposed to evaluate *ArZiGo* and Section 5 presents the obtained results. Finally, Section 6 presents the conclusions and open lines.

2. Research-paper recommender systems

The academic and scientific research world is a field where Recommender Systems (RSs) have had a great impact in the last decade. The different approaches developed in the field of RP-RSs, also known as Big Scholarly Data [7], can be classified in seven main classes:

- **Stereotyped**. Individuals are classified based on a series of characteristics. Once the stereotype is generated, the system does not need a great capacity for calculation. Most works are based on empirical studies [8] and its acceptance is quite limited [9].
- **Content-Based Filtering (CBF)**. CBF is one of the most used approaches in RP-RSs. Most of the work focuses on the representation of the content through words, phrases, n-grams, non-textual representation, or mind maps [10–12].
- **Collaborative Filtering (CF)**. The are two main characteristics which have led this technique to be tested in the area of RP-RSs: independence from context and the possibility of offering recommendations with a certain degree of serendipity. Although these systems are not so effective in domains where there are many more items than users, in recent years works with a purely collaborative approach have been developed. This is the case of [13], which implements a User-User CF approach that exploits the similarity of the user profiles. The user profiles are generated with the information the users provide when registering in the system. With respect to the feedback type, most works prefer implicit feedback, which has proved to be more useful in this field.
- **Co-occurrence**. These systems recommend publications that appear repeatedly and are related to a given article. In [14], candidate papers are obtained from both the reference papers that appear in the papers that cited a given paper of interest, and the papers that cite the papers referenced by the same paper of interest. As a final step, only the papers co-cited with the paper of interest and referenced by at least one of the papers referenced by the paper of interest are taken into account. In BibTip3[1] the co-occurrence is established between pairs of titles that are displayed together in the same session.
- **Graph-based**. The inherent connections that exist within the academic world, for example, citations and references, authors, publications or even the years of publication are used for the generation of recommendations. From these connections, graphs or networks of associations between users, between users and articles or between the articles themselves are generated. Then, graph-based algorithms can use the association graphs or networks to identify the articles that can be relevant for each user [5, 15–17].
- **Global relevance**. The simplest form of recommendation is based on the elements with the greatest global relevance or popularity. However, these kinds of recommendations are not personalized, but many approaches use this information as an additional ranking factor [5].
- **Hybrid systems**. These systems combine several techniques in different ways, from weak levels, where one of the techniques plays a dominant role, to mixed approaches, where all the techniques have similar relevance [5]. This way, they try to solve the intrinsic problems that different individual solutions have, for example cold start, data sparsity, accuracy, scalability, or diversity [18].

Most of the approaches that have been published in the related literature have hybrid characteristics, in one way or another, this also being the approach adopted in this work. Next, we present a summary of these works.

2.1. Hybrid recommender systems for scientific papers

Applied to a digital bibliography service with two million papers, [19] describes a weighted hybrid RS which combines content-based filtering and graph-based filtering. The content-based filtering is implemented in two different ways. The first one entails modeling each paper using TF-IDF and, then, computing the matrix that contains the similarities between papers. Finally, the similarities between the papers which were downloaded by the target user and the remaining papers in the catalogue are used to sort the papers. The second approach requires computing the neighborhood of the user based on the user-item matrix. The papers downloaded by the neighborhood members, but which were not downloaded by the target user, are scored considering their similarity with the papers downloaded by the user. The graph-based approach uses the citations of the papers to determine the papers to be recommended. This algorithm not only considers the papers which are cited by a specific paper, but also those papers which the cited papers refer to. The results obtained by the two previous approaches are combined, assigning different weights to each one. All the approaches are evaluated testing their precision and recall, where the hybrid system surpasses the other two.

The purpose of the work described in [20] is to recommend papers to authors with related research interests. To this end, a hybrid approach using content-based filtering and communities of researchers is presented.

Other RP-RSs rely on the construction of user profiles through the observation and interpretation of the user activities, which allows personalized recommendations to be made, and therefore user satisfaction can be increased. The work presented in [21], which follows this strategy, carries out the following two-stage recommendation method. To reduce as much as possible the number of articles to process, all the irrelevant articles are first filtered out by matching the user profile with the content of candidate articles. Next, content relevance of each

[1] http://www.bibtip.com.

candidate article, connectivity score at behavioral, social, and semantic levels and quality score that unify recency, citation and impact factor of the journal are further aggregated with the appropriate weighting distribution.

When an article is recently published, its global relevance drops due to the low number of citations. In order to avoid this issue, [22] evaluates the importance of each paper using centrality measures. The proposed method creates a multilevel citation and relationship network of authors, where a structural relationship between the papers to extract those which are significant . In the proposed methodology, a citation network is generated. Then, a candidate score is calculated by using bibliographic coupling and co-citation analysis. Next, after applying centrality measures on candidate papers, the most significant ones are identified. Finally, a relationship network of authors is extracted from significant papers to find a set of key authors and recommend the top 10 papers of authors with the highest eigenvector values.

Using as input information the keywords facilitated by the users, an algorithm which retrieves the top-k most relevant papers from a publication repository is presented in [23]. Besides the content of the paper and the citation network, their solution also uses the impact factor of venues and the vitality of papers according to the distribution of how many times each paper has been cited since it was published to the current year. In addition, a clustering method for putting papers with similar topics into one cluster based on the edge-reinforced citation network is applied, saving processing time in calculating the content-based similarity between papers and user input keywords.

A system that provides recommendations for approximately 160 million papers used by Microsoft Academic is presented in [24]. In this work, the authors combine the advantages of co-citation based and content-based approaches. The first one presumes that papers with higher co-citation counts and papers that are often cited together are more likely to be related. The main problem is that the list of references of the papers is often incomplete or unavailable. To avoid this issue, a word embedding process that subsequently combines the words to form paper embeddings is designed. Embeddings are clustered using spherical k-means to significantly reduce the number of computations. In the final step, all candidate sets are combined to create final recommendations.

An algorithm that follows a top-down methodology for filtering papers from a more generic and larger network to a community of a special interest group is developed in [25]. After the construction of the citation network, a community detection method that follows a greedy approach to optimize modularity is applied to the graph. This way, given an input paper, the algorithm identifies which community it belongs to, and narrows down the search to the nodes belonging to that particular community. After generating the network, a popularity score is assigned to each node, where a set of significant variables with impact on the number of citations is reduced to four principal components after applying the Principal Components Analysis method. Finally, the score is generated by combining these principal components.

A system that combines CBF and a CF is proposed in [26], where publicly available contextual metadata is used in order to infer the hidden associations existing between the publications.

Most of the works reviewed propose solutions that are based on the generation of networks of relationships and on the content of the articles. Although these approaches provide good results, especially in evaluations where precision and recall are measured, they contain reprehensible disadvantages such as the need for knowledge of the domain or the lack of personalization in the recommendations generated. Collaborative approaches, in addition to overcoming the aforementioned disadvantages, also provide serendipity and novelty in recommendations, characteristics that are often undervalued but highly beneficial for users.

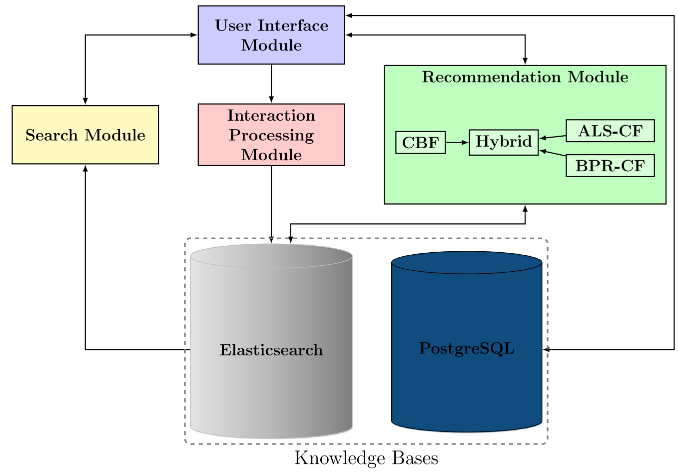

Fig. 1. Architecture of *ArZiGo*.

3. ArZiGo

ArZiGo is a system that provides personalized recommendations of scientific articles adapted to the needs and interests of the users, capable of adapting approaches ranging from citation networks, content-based and collaborative filtering thanks to its article indexing structure, its interaction processing module, and its configurable design in the recommendation module.

ArZiGo uses the Semantic Scholar Open Research Corpus (SSORC) as a corpus of articles. SSORC is a multidisciplinary article repository that contains articles associated with different scientific fields such as computer science, microbiology, neuroscience, molecular biology and biomedicine [27]. It was developed by Semantic Scholar (S2) within the Construction of the Literature Graph in the Semantic Scholar project.[2] It is a scalable system for the organization of scientific literature published in a *heterogeneous graph* to facilitate manipulation and discovery by algorithmic techniques. The corpus is continuously updated with new publications, so the scalability component must be taken into account. The corpus can be obtained as structured data that contains information such as paper abstracts, bibliographic references, citations, authors, named entities, etc.

In the following section, the architecture and the algorithms implemented in the current version of *ArZiGo* are described.

3.1. Architecture

The architecture of *ArZiGo* is composed of 5 main modules (see Fig. 1): User Interface, Knowledge Bases, Search Module, Interaction Processing Module, and Recommendation Module.

The *User Interface Module* allows the users to perform queries and to filter the results, sending the queries to the search module and displaying the results. It also provides the captured interactions to the Interaction Processing Module.

The *Search Module* provides the functionality to query the Elasticsearch knowledge base to look for articles.

The *Interaction Processing Module* is in charge of processing the interactions collected by the User Interface Module, converting them to implicit feedback and storing them in order to be used by the Recommendation Module. The main idea here is that the more the users interact with an element (article/author), the more interest they show to have in it. The information received by the module contains both the element on which the interaction is made and the type of interaction. As

[2] https://github.com/allenai/s2orc.

some interactions indicate a higher level of interest, the module assigns a different weight to each type of interaction. For each interaction with an article, besides the entry associated with the article, a new entry is generated for each named entity of the article and for each of its authors. All these entries share the same weight, which was assigned to the article interaction entry.

This module is also in charge of creating and updating the user profiles that are used by the recommender algorithms. These user profiles are based on the named entities of the content each user reads, where each named entity has a weight associated which represents the interest showed by the user.

ArZiGo uses two *Knowledge Bases* for data storage. On the one hand, the PostgreSQL relational database management system manages both the user accounts and the information of interest for the user, and, on the other hand, the Elasticsearch search engine stores the corpus of articles, user feedback, user profiles and the recommendations made.

The *Recommendation Module* is responsible for generating recommendations for the users. In the current version of *ArZiGo*, this module is in turn divided into four submodules: CBF, a memory-based recommender, which carries out recommendations considering the named entities covered by the scientific paper and the users profiles; Alternating Least Squares (ALS) and Bayesian Personalized Ranking (BPR) CF model-based recommenders, which implement different matrix factorization approaches; and finally, the *Hybrid recommender*, which follows a parallel-mixed approach that allows *ArZiGo* to integrate the other mentioned recommenders.

3.2. Algorithms

In this subsection the four recommendation algorithms implemented in the current version of *ArZiGo* are briefly presented.

- **CBF recommender**. This recommender uses the information contained in the user profile to convert it into an Elasticsearch query, where the weight values of the profile are used to specify that not all the named entity matches contribute equally to the final score. Any document matching at least one clause of the query is eligible to be returned to the recommendation module. Documents with more than one match are scored higher. The CBF recommender is launched online every time the user is validated in the system, since implemented search engine queries are fast enough. These results are indexed for further treatment.
- **ALS-CF recommender**. This model-based collaborative filtering recommender, which implements the ALS algorithm [28], is a matrix factorization approach that works by decomposing the user-item interaction matrix into the product of two lower dimensionality rectangular matrices. There are two main benefits of this approach. First, it is very easy to parallelize. Secondly, whenever dealing with implicit datasets, which are usually not sparse, it is a very efficient optimization technique. This recommender, unlike the CBF recommender, must be run offline at configurable time intervals, due to the execution time needed.
- **BRP-CF recommender**. Bayesian Personalized Ranking is another matrix factorization approach that directly optimizes its model parameters for ranking. Its optimization criterion involves pairs of items (the user-specific order of two items) to come up with more personalized rankings for each user, instead of scoring just on the user-item interaction [29]. This recommender is also run offline.
- **Hybrid recommender**. Finally, a parallel-mixed type Hybrid recommender has been implemented. As pointed out above, the recommendation module is configurable, so we can activate and deactivate each of the previous recommenders and the new ones that will be incorporated in the future. The Hybrid recommender gathers the recommendations generated by the active recommenders and normalizes and combines them to obtain the top n recommendations. The process is fast so, as in the case of the CBF Recommender, it is run online [18].

4. Evaluation

The appropriateness of the recommender item is probably the most important aspect to evaluate in a RS, in general, and in a RP-RS in particular. To evaluate the appropriateness of the recommended items, most of the works in the field keep on using the offline evaluation approach, which, despite being useful to discard algorithms in the early phases, does not seem the best way to measure the final satisfaction of users, as they can lead to erroneous conclusions regarding their effectiveness. Offline evaluation should always be supported by online evaluation or, if not possible, by rigorous user studies [30]. However, there are other aspects less taken into account, such as the scalability of the RP-RSs, which is also important [30]. To assess the scalability of the system, the computational complexity and the execution time of the techniques used must also be considered. In our case, the two model-based recommender algorithms, the ALS-CF and the BRP-CF algorithms, have been analyzed, since the other two have shown a constant behavior.

4.1. Recommendation appropriateness

Having access to adequate real data to generate and appropriately evaluate this type of system is a critical problem, since it is associated with a very long acquisition time, high economic cost and strict legal regulations associated with privacy issues [31].

The offline evaluation method is the most used method when evaluating algorithms in the RP-RSs area [30]. The offline evaluation requires the availability of a dataset which allows the items that should be recommended to be identified or the predicted ratings to be compared with the real ones. In the context of RP-RSs, there are few existing datasets that contain implicit feedback of the users. The RARD (Related-Article Recommendation Dataset), presented in [32] and subsequently updated with the RARDII version in [33], is an example. The information used to generate this dataset is collected by a Recommendations-as-a-Service (RaaS) called Mr. DLib2. However, the features of RARD do not fit with *ArZiGo*, since Mr. DLib offers recommendations based on an article provided by the system that hosts the service and therefore there is no information about users.

For the evaluation of collaborative algorithms, a broad set of user interactions is required. Both the lack of a dataset that contains implicit feedback and the fact of working with a specific corpus make the use of offline evaluations very difficult. Furthermore, the use of offline evaluations only allows the results to be analyzed from the point of view of accuracy, which can lead to misleading and biased results [34].

One of the best approaches to overcome the limitations mentioned above is the creation and use of synthetic data. In recent years, in relation to recommendation systems, the use of synthetic data is gaining more and more relevance and there are even initiatives that seek to define a methodology for the generation of synthetic data [35].

The generation of synthetic data can be carried out at different levels, since the needs of the systems are different. These range from protecting sensitive data by applying different transformations or substitutions [36], to the creation of complete synthetic datasets [37, 38].

Next, the evaluation methodology conducted on *ArZiGo* is presented.

4.2. Methodology

In this work, a synthetic data generator has been developed as a solution to create a suitable dataset for the evaluation of *ArZiGo*. This data generator relies on predefined profiles, and generates a random number of users for each profile along with the sessions and their interactions with *ArZiGo*. This dataset simulates the data *ArZiGo* would gather in a real scenario.

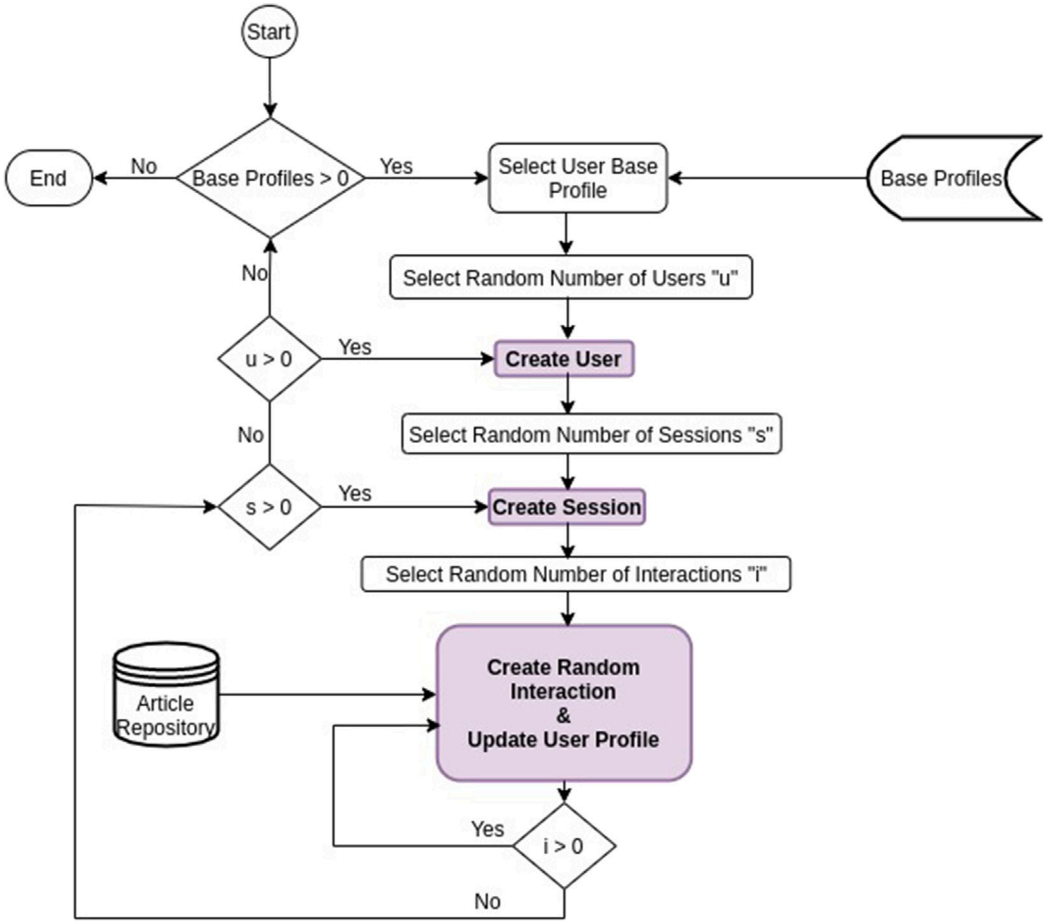

Fig. 2. Synthetic data generator.

In the evaluation, a reduced version of the SSORC repository containing 23 million articles has been extracted. Regarding the research areas, 10 profile groups or domains have been chosen, 7 of which belong to the area of *Computer Science* and 3 of which belong to the area of *Medicine*: *Computer vision, Image processing, Sentiment analysis, Word sense disambiguation, Question answering, Machine Translation, Recommender Systems, Alzheimer's disease, Sudden cardiac death*, and *Lung disease*. This way, the performance of the CF models can be evaluated at different levels. On the one hand, with users whose profiles are far from each other, and, on the other hand, with users whose profiles are close to each other, as in the case of the NLP domain.

To ensure the correct evaluation of the recommendations, a group of experts from the different domains have participated in the evaluation process, specifically 30 experts.

Next, the evaluation methodology for *ArZiGo* is detailed:

1. Synthetic data generation.
2. Pseudo-random selection of users for evaluation.
3. Generation of recommendations for the selected users.
4. Expert evaluation of the generated recommendations.

Fig. 2 illustrates the process that generates users and their interactions. As a result, 684 users have been created with a total number of 3,500,000 interactions approximately.

Once the users have been created, a pseudo-random selection of the users whose recommendations will be evaluated has been carried out, where the selected number has been limited to the number of experts recruited by each domain. To avoid any kind of bias, users with a profile which is too general have been discarded.

Each human expert receives a form with the recommendations to evaluate, which includes the following data:

- A brief description of the file content and the instructions to complete the evaluation.
- The profile information with the named entities that comprise it, along with the percentage weight of each one.
- Information about the recommended articles.

In the evaluation process carried out in this work, a group of 30 human experts participated in the evaluation of 500 recommendations belonging to 10 different research areas. The expert who conducted the evaluation of the user's recommendations received a form with the suggested 10 articles, including the abstract and the link to the pdf documents, and had to indicate whether or not the recommendation was appropriate. In addition, they could include an additional comment explaining their decision.

5. Results and discussion

In this section, besides the outcomes of *ArZiGo*, the performance of the model-based algorithms is also analyzed.

5.1. Analysis of the appropriateness of the recommendation

In order to measure the results of the evaluations, there are two metrics that have been considered, the *precision* and the *mean reciprocal rank* (MRR). The precision, or fraction of relevant instances among the retrieved articles, is measured by the opinion of the experts about the correctness of the given recommendations. The MRR metric, on the other hand, measures the position of the first correct recommendation. Both measures have been analyzed from different points of view.

Table 1
Recommendations and *precision* results per domain.

Domain	# Experts	# Recom.	Correct	Incorrect	MRR
Computer vision	2	50	60%	40%	54%
Image processing	2	50	88%	12%	80%
Sentiment analysis	3	50	82%	18%	86%
Word sense disambiguation	3	50	32%	68%	44%
Question answering	3	50	54%	46%	76%
Machine translation	3	50	40%	60%	43%
Recommender Systems	3	50	100%	0%	100%
Alzheimer's disease	2	50	76%	24%	76%
Sudden cardiac death	2	50	85%	15%	90%
Lung disease	2	50	72%	28%	100%
TOTAL	**30**	**500**	**69%**	**31%**	**75%**

Table 2
Precision results per recommender.

Recommender	# Recommendations	Correct	Incorrect
CBF	116	69%	31%
ALS-CF	382	70%	30%
BPR-CF	395	69%	31%

Table 1 shows the *precision* achieved by *ArZiGo* for every domain as well as the number of experts recruited and the number of recommendations reviewed. It can be observed that the correct results range from 100% for *Recommender Systems* to 32% for *Word sense disambiguation*.

The worst results are obtained in three NPL sub-domains, *Word sense disambiguation*, *Question answering* and *Machine translation*.

While Table 1 displays the outcomes of the Hybrid Recommender of *ArZiGo*, Table 2 shows the performance of the recommenders *ArZiGo* integrates. As can be observed, the three algorithms obtain similar results, even though the contribution of the CBF recommender to the final recommendations is less. It must be taken into account that a recommendation could have been provided by one or more recommenders. The data in the table takes into account all the contributions of each recommender.

Finally, Table 3 shows the *precision* results per domain and recommender.

In addition to knowing which of the proposed recommendations are correct, we are also interested in measuring their position in the list of recommendations, since the objective is to offer a list of recommendations ordered by interest of the user. To measure this, the *MRR* has been selected.

The *MRR* metric is calculated by applying the following equation:

$$MRR = \frac{1}{N}\sum_{n=1}^{N}\frac{1}{rank_i}, \tag{1}$$

where $rank_i$ indicates the position of the first relevant recommendation made and N the number of users or profiles.

Both the results obtained per domain and the final result for this measure are shown in Table 1.

The final result of 0.75 indicates that more than half of the profiles receive the first relevant recommendation in the first position. Again, the poorest results are obtained in NLP-related domains.

In order to delve into these results, the percentages of the first correct recommendation per classifier have been extracted. The obtained results (see Table 4) again show a very similar behavior between the three recommenders.

5.2. Analysis of the execution time

As pointed out above, conducting an analysis of the response times of the implemented model-based algorithms, the ALS-CF recommender and the BRP-CF algorithm, has also been considered relevant.

Since *ArZiGo* is a scalable system, it is important to analyze the performance of the implemented algorithms in different scenarios, so, in addition to the dataset generated to obtain the recommendations for the evaluation process, two more datasets of different sizes have been created. In Table 5, the features of these two additional datasets are described.

Fig. 3, shows the time in seconds needed to generate each model. It can be observed that the most time consuming algorithm is *ALS-CF*. On the other hand, the *ALS-CF* algorithm is the one that suffers the most when increasing the size of the dataset, while the time spent in generating the models for *BPR-CF* remains practically uniform.

Fig. 4 shows the time in seconds needed to generate the recommendations per model. In this case, *ALS-CF* is again the one that requires most time to generate recommendations. Except for the peak in *ALS-CF* for *dataset_M*, the slope is quite soft, which seems to show that the requirement for efficiency and scalability has been satisfied.

6. Conclusions and future work

This paper has presented *ArZiGo*, a web-based system for the recommendation of scientific articles that manages a continuously growing multidisciplinary scientific literature repository, namely, the *Semantic Scholar Open Research Corpus*, making personalized recommendations of research articles to users. *ArZiGo* has been developed to fulfill the needs of researchers and students who are introducing into research activities. Besides allowing researchers to search for articles they are interested in and managing the discovered references, *ArZiGo* recommends the users papers according to their interests. Furthermore, the results of the searches are ranked according to the profiles of the users to facilitate the identification of the papers they are interested in. The more information in the profile of the user, the more accurate the ranking. Different approaches and recommendation techniques identified in the revision of the state-of-the-art are supported and integrated into a modular, easily extensible and open source based architecture. At present, the current prototype of *ArZiGo* integrates a hybrid research paper recommender system that combines a content-based filtering algorithm together with two collaborative filtering algorithms. In any case, integrating a new technique in *ArZiGo* is quite straightforward, due to its modular architecture.

The evaluation of *ArZiGo* has required the development of a synthetic data generator capable of generating the interactions of the users with the research papers of their interest.[3] The results obtained are quite satisfactory and this encourages us to penetrate further into the lines that are open.

The main objective is to continue researching, adapting and developing new recommendation models in the RP-RSs domain, allowing us to work with more enriched user profiles in order to get a higher level of personalization in the recommendations. For example, it is worth delving into aspects such as the difference between novice users and experts or students with more ephemeral interests and researchers in longer-term activities. Time and sequence-aware techniques, such as the algorithm proposed by Sánchez and Bellogín [39], can be adapted to recommend scientific papers to master or Ph.D. students. The hybrid module could play an important role by selecting the most appropriate approach in each case. The broader the range of characteristics considered to make the recommendation, the lesser will be the risk of insulating users from exposure to different academic viewpoints [40].

Although researchers do not rate the publications they read, they often add comments or reviews in the reference managers they use. Taking this into account, *ArZiGo* also allows the users to add comments to the publications they add to their personal library. These reviews could be a reliable means of identifying the opinions of the users about the papers [41], and, therefore, *ArZiGo* might integrate this information to better infer the interests of the users.

[3] The dataset used for the evaluation is publicly available for the research community at https://github.com/IratxePinedoehu/arzigo.

Table 3
Precision results per recommender and domain.

Domain	Recommender	# Recommendations	Correct	Incorrect
Computer vision	CBF	15	65%	35%
	ALS-CF	29	58%	42%
	BPR-CF	38	55%	45%
Image processing	CBF	8	92%	8%
	ALS-CF	41	88%	12%
	BPR-CF	41	90%	10%
Sentiment analysis	CBF	14	86%	14%
	ALS-CF	38	81%	19%
	BPR-CF	36	82%	18%
Word sense disambiguation	CBF	10	60%	40%
	ALS-CF	39	26%	74%
	BPR-CF	38	26%	74%
Question answering	CBF	11	45%	55%
	ALS-CF	42	67%	33%
	BPR-CF	41	63%	37%
Machine translation	CBF	9	11%	89%
	ALS-CF	43	51%	49%
	BPR-CF	43	51%	49%
Recommender Systems	CBF	5	100%	0%
	ALS-CF	47	100%	0%
	BPR-CF	48	100%	0%
Alzheimer's disease	CBF	11	64%	36%
	ALS-CF	40	76%	24%
	BPR-CF	39	77%	23%
Sudden cardiac death	CBF	17	79%	21%
	ALS-CF	39	92%	8%
	BPR-CF	40	91%	9%
Lung disease	CBF	18	89%	11%
	ALS-CF	24	67%	33%
	BPR-CF	29	66%	34%

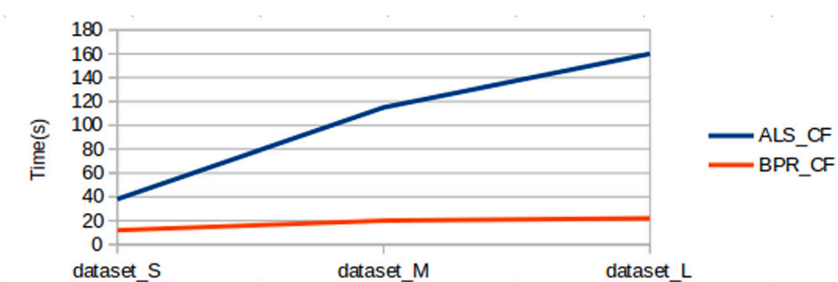

Fig. 3. Execution time in the generation of the models.

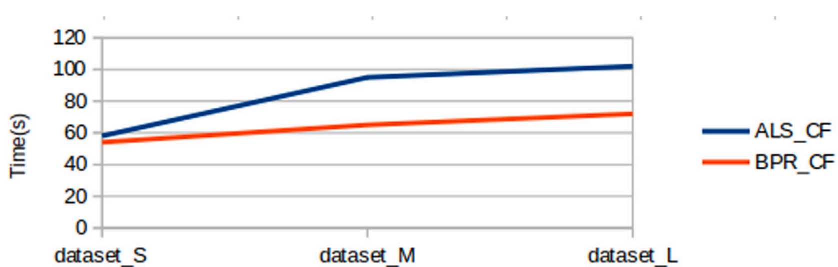

Fig. 4. Execution times for recommendation generation and their indexing.

Table 4
Results per recommender for the first correct recommendation.

CBF	ALS-CF	BPR-CF
60%	62%	62%

Table 5
Generated datasets for the purpose of comparing execution times.

Index	Interactions	Users
dataset_L	3,500,000	684
dataset_M	1,500,000	538
dataset_S	650,000	422

Another interesting research line would be the exploration of more efficient alternatives to conduct the evaluation. One of the aspects to be tackled is related to the evaluation of these kinds of systems. Although user studies might be very reliable, they are slow and require multiple resources.

Funding

This work is supported by RTI2018-096846-B-C21 (MCIU/AEI/FEDER, UE – Spanish Ministry of Science, Innovation and University/Spanish State Research Agency/European Regional Development Fund) and ADIAN grant IT-1437-22 (Basque Government).

Declaration of competing interest

We do not have conflict of interests.

Data availability

No data was used for the research described in the article.

References

[1] P.K. Singh, P.K.D. Pramanik, A.K. Dey, P. Choudhury, Recommender systems: an overview, research trends, and future directions, Int. J. Bus. Syst. Res. 15 (1) (2021) 14–52, URL https://EconPapers.repec.org/RePEc:ids:ijbsre:v:15:y:2021:i:1:p:14-52.
[2] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: A survey, Decis. Support Syst. 74 (2015).
[3] T. Pradhan, S. Pal, Cnaver: A content and network-based academic venue recommender system, Knowl.-Based Syst. 189 (2020) 105092.
[4] T. Pradhan, A. Gupta, S. Pal, Hasvrec: A modularized hierarchical attention-based scholarly venue recommender system, Knowl.-Based Syst. 204 (2020) 106181.
[5] J. Beel, B. Gipp, S. Langer, C. Breitinger, Research-paper recommender systems: A literature survey, Int. J. Digit. Libr. (2015) 1–34.
[6] C. Kreutz, R. Schenkel, Scientific paper recommendation systems: a literature review of recent publications, Int. J. Digit. Libr. 23 (2022) 1–35.
[7] S. Khan, X. Liu, K. Shakil, M. Alam, A survey on scholarly data: From big data perspective, Inf. Process. Manage. 53 (2017) 923–944.
[8] N.A. ALRossais, D. Kudenko, Evaluating stereotype and non-stereotype recommender systems, in: KaRS@RecSys, 2018.
[9] J. Beel, S. Dinesh, P. Mayr, Z. Carevic, R. Jain, Stereotype and most-popular recommendations in the digital library sowiport, in: ISI, 2017.
[10] X. Bai, M. Wang, I. Lee, Z. Yang, X. Kong, F. Xia, Scientific paper recommendation: A survey, IEEE Access 7 (2019) 9324–9339.
[11] M. Passon, M. Comuzzo, G. Serra, C. Tasso, Keyphrase extraction via an attentive model, in: Proceedings of the 15th Italian Research Conference on Digital Libraries, IRCDL 2019, 2019, pp. 304–314.
[12] K. Bennani-Smires, C. Musat, A. Hossmann, M. Baeriswyl, M. Jaggi, Simple unsupervised keyphrase extraction using sentence embeddings, in: Proceedings of the 22nd Conference on Computational Natural Language Learning, CoNLL, 2018.
[13] B. Madhushree, A novel research paper recommendation system, Int. J. Adv. Res. Eng. Technol. (IJARET) 7 (2016) 7–16.
[14] N. Sakib, R.B. Ahmad, K. Haruna, A collaborative approach toward scientific paper recommendation using citation context, IEEE Access 8 (2020).
[15] X. Ma, R. Wang, Personalized scientific paper recommendation based on heterogeneous graph representation, IEEE Access 7 (2019).
[16] M. G, A. P, H. MR, M. G, M. GG, Cold start problem alleviation in a research paper recommendation system using the random walk approach on a heterogeneous user-paper graph, Int. J. Intell. Inf. Technol. 16 (2020) 24–48.
[17] W. Tanner, E. Akbas, M. Hasan, Paper recommendation based on citation relation, 2019, pp. 3053–3059.
[18] E. Çano, M. Morisio, Hybrid recommender systems: A systematic literature review, Intell. Data Anal. 21 (2017) 1487–1524.
[19] Y.-C. Lee, J. Yeom, K. Song, J. Ha, K. Lee, J. Yeo, S.-W. Kim, Recommendation of research papers in dbpia: A hybrid approach exploiting content and collaborative data, in: 2016 IEEE International Conference on Systems, Man, and Cybernetics, SMC, 2016, pp. 002966–002971.
[20] A. Tsolakidis, E. Triperina, C. Sgouropoulou, N. Christidis, Research publication recommendation system based on a hybrid approach, in: Proceedings of the 20th Pan-Hellenic Conference on Informatics, 2016, pp. 1–6.
[21] J. Sun, Y. Jiang, X. Cheng, W. Du, Y. Liu, J. Ma, A hybrid approach for article recommendation in research social networks, J. Inf. Sci. 44 (5) (2018) 696–711.
[22] W. Waheed, M. Imran, B. Raza, A.K. Malik, H.A. Khattak, A hybrid approach toward research paper recommendation using centrality measures and author ranking, IEEE Access 7 (2019).
[23] Q. Yang, Z. Li, A. Liu, G. Liu, L. Zhao, X. Zhang, M. Zhang, X. Zhou, A Novel Hybrid Publication Recommendation System Using Compound Information, Vol. 22, World Wide Web, 2019.
[24] A. Kanakia, D. Eide, Z. Shen, K. Wang, A scalable hybrid research paper recommender system for microsoft academic, in: The Web Conference 2019 - Proceedings of the World Wide Web Conference, WWW 2019, 2019.
[25] P. Yadav, N. Remala, N. Pervin, Reccite: A hybrid approach to recommend potential papers, in: 2019 IEEE International Conference on Big Data, Big Data, 2019, pp. 2956–2964.
[26] K. Haruna, M.A. Ismail, A. Qazi, H.A. Kakudi, M. Hassan, S.A. Muaz, H. Chiroma, Research paper recommender system based on public contextual metadata, Scientometrics 125 (2020).
[27] W. Ammar, D. Groeneveld, C. Bhagavatula, I. Beltagy, M. Crawford, D. Downey, J. Dunkelberger, S. Feldman, V. Ha, R. Kinney, S. Kohlmeier, K. Lo, T. Murray, H.-H. Ooi, M. Peters, J. Power, S. Skjonsberg, L. Wang, oren, Construction of the literature graph in semantic scholar, 2018, pp. 84–91.
[28] Y. Hu, Y. Koren, C. Volinsky, Proceedings - IEEE International Conference on Data Mining, ICDM, 2008, pp. 263–272.
[29] S. Rendle, C. Freudenthaler, Z. Gantner, L. Schmidt-Thieme, Bpr: Bayesian personalized ranking from implicit feedback, in: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence, UAI 2009, 2012.
[30] J. Beel, S. Langer, A comparison of offline evaluations, online evaluations, and user studies in the context of research-paper recommender systems, in: 19th International Conference on Theory and Practice of Digital Libraries, TPDL, Vol. 9316, 2015, pp. 153–168.
[31] S.D.P. Mendonça, Y.P.D.S. Brito, C.G.R.D. Santos, R.D.A.D. Lima, T.D.O.D. Araújo, B.S. Meiguins, Synthetic datasets generator for testing information visualization and machine learning techniques and tools, IEEE Access 8 (2020) 82917–82928.
[32] J. Beel, Z. Carevic, J. Schaible, G. Neusch, Rard: The related-article recommendation dataset, D-Lib Mag. 23 (2017).
[33] J. Beel, B. Smyth, A. Collins, Rard II: The 2nd related-article recommendation dataset, 2018, ArXiv abs/1807.06918.
[34] J. Beel, S. Langer, M. Genzmehr, B. Gipp, A. Nürnberger, A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation, in: ACM International Conference Proceeding Series, 2013.
[35] M.D. Ekstrand, A. Chaney, P. Castells, R. Burke, D. Rohde, M. Slokom, Simurec: Workshop on synthetic data and simulation methods for recommender systems research, in: Fifteenth ACM Conference on Recommender Systems, RecSys ’21, Association for Computing Machinery, New York, NY, USA, 2021, pp. 803–805.
[36] M. Slokom, Comparing recommender systems using synthetic data, in: Proceedings of the 12th ACM Conference on Recommender Systems, RecSys ’18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 548–552.
[37] M. Pasinato, C.E. Mello, M.-A. Aufaure, G. Zimbrão, Generating synthetic data for context-aware recommender systems, in: 2013 BRICS Congress on Computational Intelligence and 11th Brazilian Congress on Computational Intelligence, 2013, pp. 563–567.
[38] M. del Carmen Rodríguez-Hernández, S. Ilarri, R. Hermoso, R. Trillo-Lado, Datagencars: A generator of synthetic data for the evaluation of context-aware recommendation systems, Pervasive Mob. Comput. 38 (2017) 516–541, special Issue IEEE International Conference on Pervasive Computing and Communications (PerCom) 2016.
[39] P. Sánchez, A. Bellogín, Time and sequence awareness in similarity metrics for recommendation, Inf. Process. Manage. 57 (3) (2020) 102228, URL https://www.sciencedirect.com/science/article/pii/S0306457319307678.
[40] A. Polonioli, The ethics of scientific recommender systems, Scientometrics 126 (2) (2021) 1841–1848, URL https://ideas.repec.org/a/spr/scient/v126y2021i2d10.1007_s11192-020-03766-1.html.
[41] D. Margaris, C. Vassilakis, D. Spiliotopoulos, What makes a review a reliable rating in recommender systems? Inf. Process. Manage. 57 (6) (2020) 102304, URL https://www.sciencedirect.com/science/article/pii/S0306457320307998.